

21E CIELE de Educación Estocástica

EXPLORACIÓN DINÁMICA DE DATOS CON TINKER PLOTS

Rivera Flórez, Tulia Esther
trivera@uis.edu.co
Universidad Industrial de Santander

RESUMEN

Tinker plots es un software educativo para exploración dinámica de datos. Su filosofía difiere de los programas convencionales en que el objetivo es que el estudiante vaya construyendo progresivamente gráficos que le permiten explorar los datos, justificar conclusiones y formular sus propias hipótesis. Entre sus potencialidades se destaca su interfaz que ofrece opciones de interacción novedosas que permiten una fácil manipulación de los datos, sumadas a herramientas gráficas para modelamiento y simulación. El taller tiene como propósito hacer una descripción general de los recursos y características del software y permitir la discusión sobre las opciones didácticas que ofrece el mismo.

PALABRAS CLAVE

Análisis exploratorio, Software estadístico, Tinker plots, TIC en el aula.

INTRODUCCIÓN

Como ciudadanos de la era de la información se está cada vez más expuesto a tener que interactuar con información basada en datos. Bajo esta perspectiva, un foco importante de la formación es aprender a razonar con datos acorde al enfoque que plantea la American Statistical Association (2014) en su Guía curricular, donde se sugiere hacer del currículo en Estadística una integración entre teoría, métodos, software y aplicaciones haciendo uso de metodologías que conduzcan a enfatizar en la importancia del trabajo con datos, utilizar software que permita manipular y gestionar datos en varias formas e incentivar el razonamiento intuitivo.

También, hoy en día es posible ver como la denominada Ciencia de los datos se considera parte de la Estadística moderna; actualmente dicha área tiene como prioridad el estudio y desarrollo de herramientas para analizar grandes bases de datos. Ante este panorama, surge la necesidad de involucrar en el aula de clase recursos que posibiliten el desarrollo de habilidades básicas del pensamiento estadístico que brinden soporte al proceso de toma de decisiones basados en datos. El objetivo de este taller es mostrar un software educativo desarrollado bajo esta directriz.

MARCO DE REFERENCIA

ANÁLISIS EXPLORATORIO DE DATOS

Enfoque alternativo para el análisis de datos propuesto por Batanero, Estepa y Godino (1991) quienes lo definen como:

El estudio de los datos desde todas las perspectivas y con todas las herramientas posibles, incluso las ya existentes. El propósito es extraer cuanta información sea posible, generar hipótesis nuevas, en el sentido de conjeturar sobre las observaciones de las que disponemos (p. 2).

El contraste con el análisis de datos clásico es que no se parte de un modelo o hipótesis preconcebida, el objetivo es usar información gráfica para conocer ampliamente los datos antes de pasar a la etapa de cálculos, procesos de prueba de hipótesis o ajuste de modelos. Como punto de referencia para situar el trabajo bajo este enfoque puede citarse el uso preferencial que hace de la mediana y las medidas de posición como descriptores básicos sobre la tradicional dupla, promedio y desviación estándar, medidas que son adecuadas en presencia de normalidad.

TINKER PLOTS

Software educativo creado por Clifford Konold y Craig Miller en el marco del proyecto *Developing tools and curricula for enhancing data analysis in the middle school*. Toda la información relacionada con el programa se puede consultar en <http://www.srri.umass.edu/tinkerplots-project>.

Es un software para visualización de datos y modelamiento, concebido como herramienta de apoyo a los procesos de formación dirigido a un público entre los 9 y 13 años de edad. Su potencial radica en las posibilidades que ofrece para la visualización y modelamiento de datos. Aparece en 2004 tras cinco años en fase de desarrollo. A diferencia de la mayoría de programas estadísticos que ofrecen comandos para construir gráficos dispuestos en una barra de menús, Tinker plots está pensando para que sea el estudiante quien construya los gráficos en forma progresiva a partir de operaciones familiares como son: apilar, ordenar y separar. Otro aspecto por resaltar de este software es que está alineado con los estándares curriculares del NCTM y que además sus creadores tuvieron entre sus prioridades dar respuesta a tres interrogantes básicos (Konold, 2007):

1. ¿Cuáles son las ideas claves en Estadística y Análisis de datos?
2. ¿Cuáles son las capacidades estadísticas que necesita un ciudadano de hoy? y ¿qué necesitarán ellos en 25 años?
3. ¿Cómo empezar a desarrollar esa capacidad en edades tempranas?

La interfaz de trabajo contiene los elementos típicos del ambiente Windows pero admite interactuar con iconos en diferentes posiciones de la pantalla y manipular las variables y los registros en forma similar a cuando se barajan cartas sobre una mesa. Además del software, los creadores han desarrollado tutoriales para familiarizar al usuario tanto con el manejo del

programa como para la enseñanza de temas de específicos (Konold & Dever, 2007) este material ha sido el insumo para este taller.

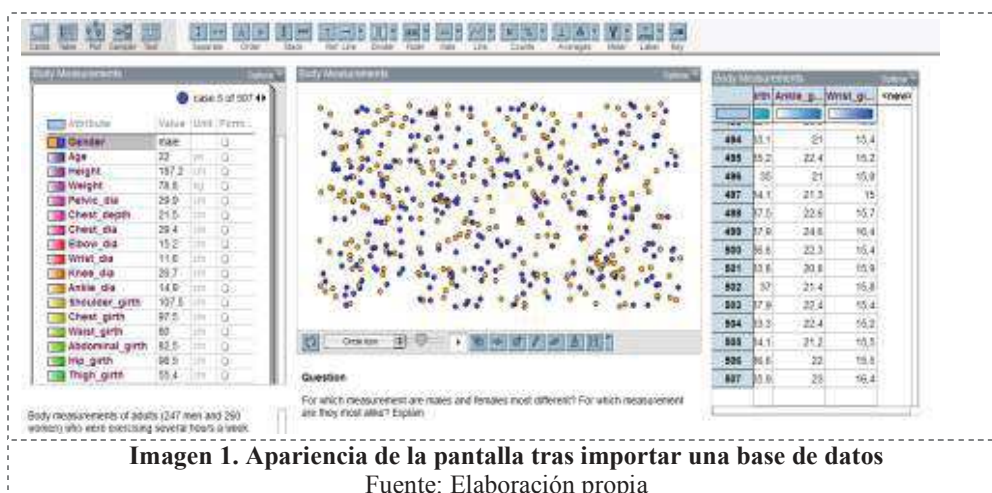
DESCRIPCIÓN DE LAS ACTIVIDADES

El taller se basa en el uso del programa Tinker plots. Las actividades incluyen la presentación de los principales objetos y comandos del programa, una actividad práctica con datos y al final el planteamiento de una pregunta que permita la puesta en común de ideas sobre posibilidades que ofrece el programa como mediador en la enseñanza de la Estadística.

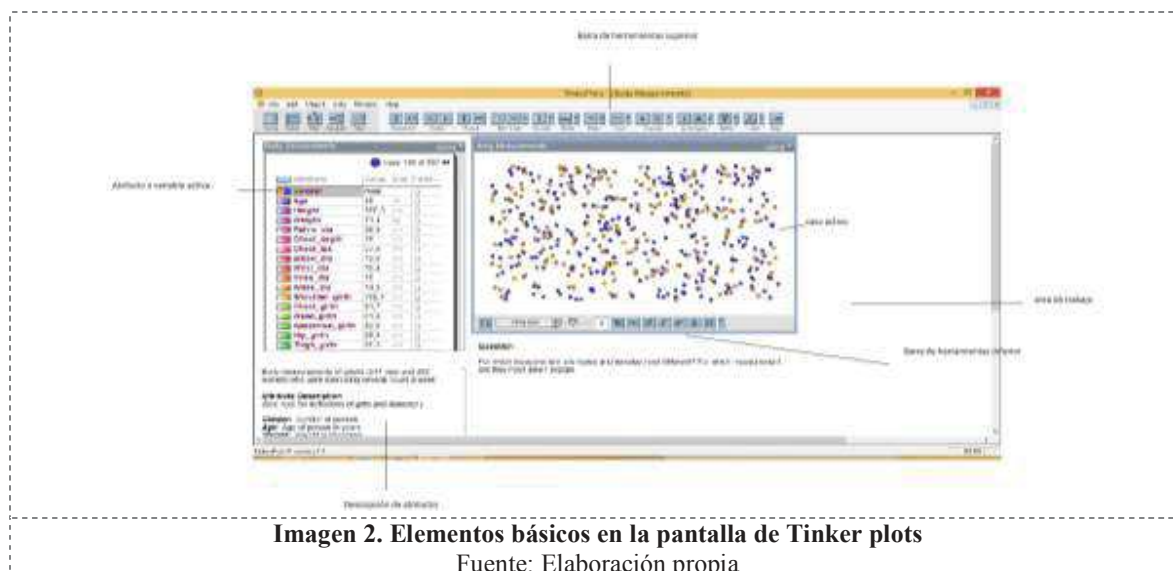
ACTIVIDAD 1: INTERFAZ Y DICCIONARIO BÁSICO DE TINKER PLOTS

- Objetivo: Revisar los elementos básicos de la pantalla, describir los objetos básicos del programa, recursos y definición de términos.

Luego de la presentación alusiva a la historia del programa, se inicia por cargar una de las bases de datos que trae el programa, “*Body measurements.tp*”, siguiendo los comandos *File – Open Sample Document - Data and Demos – Health*. Luego se hará una descripción de los elementos de la pantalla tal como se muestra en las imágenes a continuación.



Inicialmente sólo aparecen las ventanas de casos y gráficos pero como se muestra se puede activar la Tabla de datos dando clic en Table. Como se ve las variables y los casos todo el tiempo están disponibles sobre el escritorio de trabajo.



DEFINICIONES BÁSICAS

- **Colección:** base de datos, se guardan como un archivo con la extensión tp.
- **Caso:** es la unidad básica de una colección, corresponde a lo que se llama registro en una base de datos.
- **Atributo:** los casos tienen atributos, corresponden a las variables. Aparecen listadas a la izquierda y se repiten en cada tarjeta de casos. Pueden ser de tipo numérico o categórico. El icono que acompaña cada atributo va coloreado en escala descendente de un mismo color si es numérico o en diferentes colores si es categórico.
- **Valor:** cada caso tiene un valor para cada atributo.
- **Bins:** son divisiones en los gráficos, el término podría traducirse como categorías o como intervalos para el caso de atributos numéricos. Se activan arrastrando la variable desde la ventana de casos hacia el área del gráfico bien sea al eje horizontal o vertical. Se presentan algunos ejemplos en la Imagen 3.

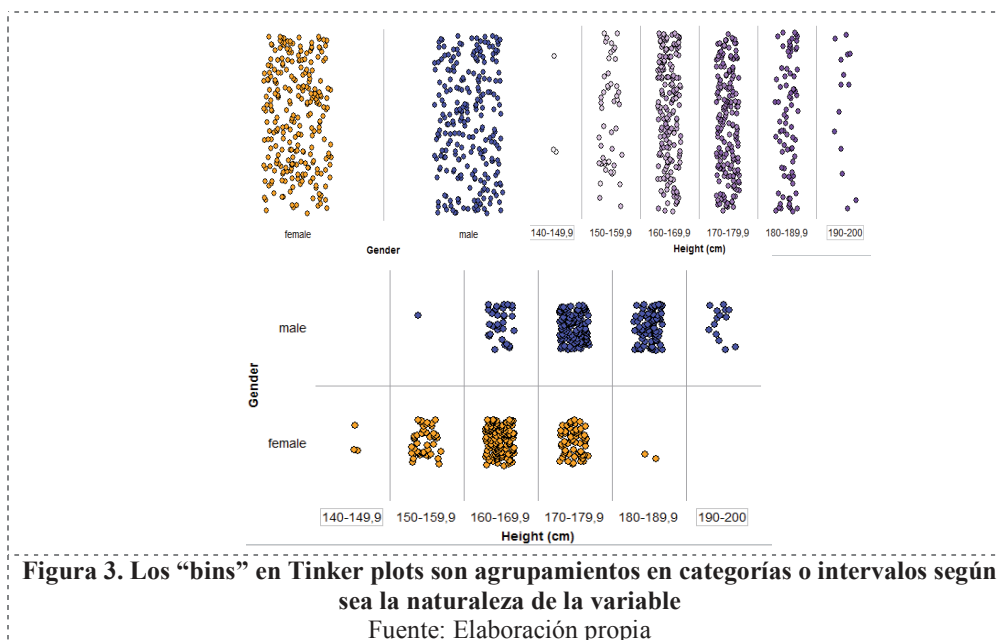


Figura 3. Los “bins” en Tinker plots son agrupamientos en categorías o intervalos según sea la naturaleza de la variable

Fuente: Elaboración propia

- Total separación de atributos: cuando se separa un atributo (variable) en todas las categorías posibles; si la variable es numérica, el eje aparecerá con una escala continua, no aparecerán líneas de división. Por ejemplo, para los datos de la variable edad, se debe tomar cualquier punto y arrastrarlo hacia la derecha hasta conseguir una apariencia como la que figura a continuación:

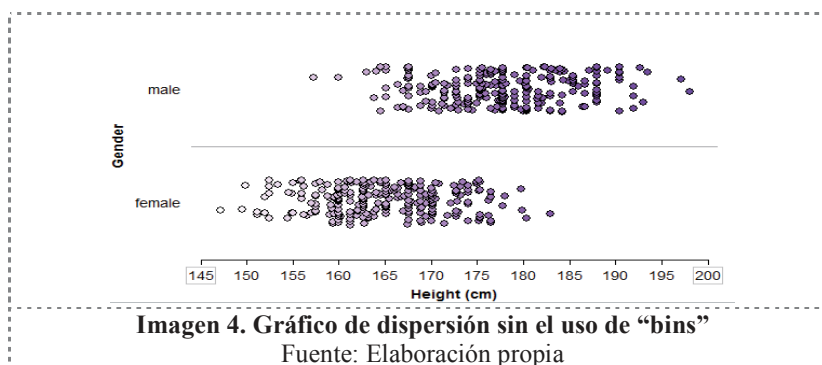


Imagen 4. Gráfico de dispersión sin el uso de “bins”

Fuente: Elaboración propia

- Casos excluidos: aparecen coloreados en gris y se ubican al lado derecho del gráfico.
- Mezclar: como su nombre lo indica revuelve todos los casos deshaciendo las acciones implementadas en el gráfico. Al activarlo, volvemos a la presentación inicial de los casos. Corresponde al ícono de la barra de herramientas inferior.
- Tipo del ícono para los casos: hay varios tipos de símbolos para representar los casos: círculos, cuadrados, barras, una detonación rectangular o imágenes. Usar los botones y el slider para cambiar el tipo y tamaño de símbolos en el gráfico. Observe que allí es posible también activar gráficos circulares y de barras.

- Fórmulas con los atributos: se pueden crear nuevos atributos usando fórmulas. Si por ejemplo se desea tener la estatura medida en metros necesitamos dar clic derecho sobre la ventana de casos, del menú que se despliega seleccionar *New attribute* y digitar el nombre que en este caso puede ser *EstMetros*; luego, dar clic derecho otra vez en la Ventana de casos para elegir *Edit Formula* que permitirá introducir un cálculo tal como se indica a continuación. También puede acceder a los comandos anteriores a través del menú Edit en la barra de menús.



Luego, se explorarán los diferentes comandos disponibles en la barra de herramientas superior para ver el efecto en la gráfica.



Pregunta: ¿Qué comandos o características encuentra en Tinker plot que lo diferencian de otros programas educativos? ¿Cuáles comandos o características podrían contribuir a superar problemas frecuentes a la hora de analizar datos?

CONCLUSIONES

Desde el punto de vista de la investigación en Educación Estadística, los autores del programa reportan que el software ha sido utilizado como recurso en trabajos sobre construcción de razonamiento inferencial informal, solución de problemas usando medidas de tendencia central y variación, creación e interpretación de gráficos estadísticos, razonamiento con muestras y muestreo, exploración de ideas intuitivas sobre aleatoriedad y caracterización de estrategias de trabajo en Estadística. En este caso, se ha valorado los recursos que ofrece el programa de manera exploratoria, encontrando que para los niños su interfaz resulta ser muy atractiva y motivadora; en cuanto a objetivos académicos, el programa resulta ser muy eficiente en cuanto al tiempo y logro en procesos de investigación basada en gráficos. Se espera tener resultados formales más adelante.

REFERENCIAS

- American Statistical Association. (2002). Curriculum guidelines for undergraduate programs in statistical science. Retrieved March 2, 2005.
- Konold, C. (2007), *Designing a data tool for learners*, In M. Lovett & P. Shah (Eds.), *Thinking with data: The 33rd Annual Carnegie Symposium on Cognition*, pp. 267--292. Hillside, NJ: Lawrence Erlbaum Associates.
- Konold, C. & Dever, H. (2007). *Thinker plots workshop guides*. California, Estados Unidos. Key curriculum press.
- Batanero, C., Estepa, A., & Godino, J. D. (1991). *Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria*. Suma, 9, 25-31.

ANEXO 1: ACTIVIDADES

Actividad 2: Analizar datos usando las herramientas de Tinker plots.

Objetivo: Usar las herramientas gráficas y de medida para responder una pregunta de investigación.

Considerando otra vez la base de datos “*Body measurements.tp*”, se plantea al grupo la pregunta: ¿En cuáles atributos (variables) hay más diferencias por género?

El resultado esperado es presentar gráficos que permitan responder a la pregunta de investigación haciendo uso de las herramientas de presentación y de medida disponibles (Ver Figura 6). Durante el análisis se enfatizarán dos de los elementos novedosos que aporta el programa: la posibilidad de analizar covariación en atributos numéricos a partir de la escala de colores y la descripción basada en el “Hat plot”, una nueva versión del gráfico de caja y bigotes.

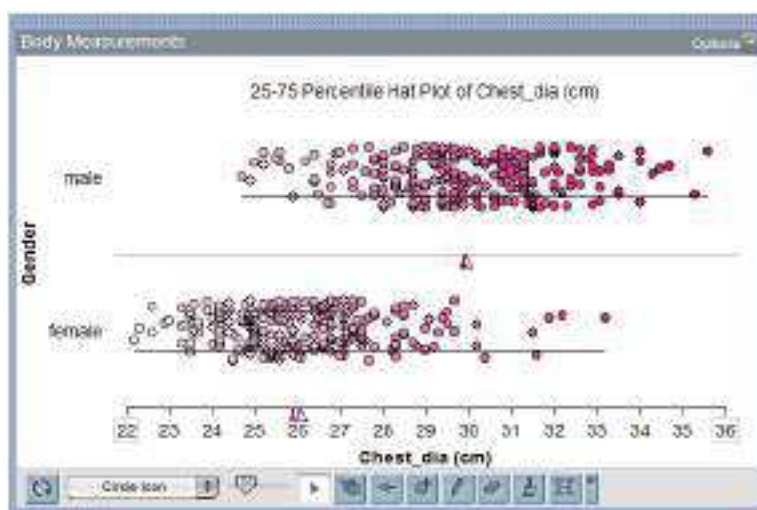


Figura 5: Gráfico elaborado en Tinker plots. Se muestra además de una segmentación de los datos basada en el género, la media (triángulo azul), la mediana (marca roja) y el hat plot (superpuesto a los datos en negro).

Pregunta: ¿Cómo implementar una sesión de trabajo orientada a la investigación basada en datos?, ¿cuál es el papel del profesor cuando se implementan este tipo de actividades?

Actividad 3: Simulando datos en Tinker plots.

Objetivo: Explorar la creación de una Fábrica de datos, es decir de la creación de escenarios para hacer simulación.

Para construir una Fábrica de datos arrastre el botón “Sampler” de la barra de menús al área de trabajo. Se agregará una ventana como la que se muestra a continuación. En la parte inferior de esta ventana aparecen los mecanismos de selección disponibles:

Mixer: selecciona elementos de un conjunto de valores discreto.

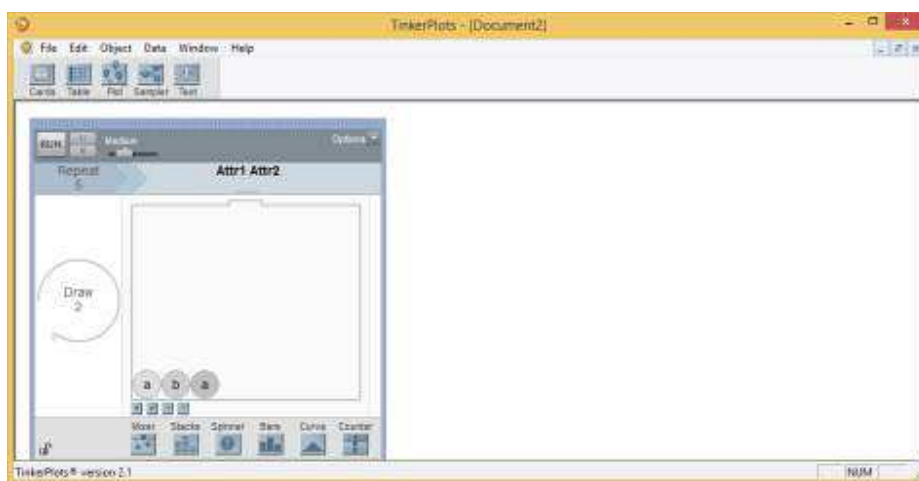
Stacks: selecciona elementos de un conjunto de valores discreto, se utiliza cuando hay valores repetidos.

Spinner: selecciona elementos de un conjunto de valores discreto, los valores pueden tener diferente probabilidad de ocurrir acorde a la distribución en la ruleta.

Bars: hace lo mismo que Spinner pero la distribución de probabilidades se determina a partir de un diagrama de barras equiprobable donde se arrastra el cursor para ir borrando.

Curve: permite modelar una distribución de probabilidad continua usando el mouse.

Counter: selecciona valores sistemáticamente



Para ilustrar su uso se construirá el mecanismo de selección Cat Factory¹ que se muestra a continuación. La máquina permite seleccionar 5 casos aleatorios con mediciones para los atributos: Nombre, Género, Color de ojos y Longitud. Cada vez que se presiona el botón RUN se selecciona una nueva muestra.

¹ Tutorial de Tinker plots disponible en <http://www.tinkerplots.com/tutorials/simulating-data>